Is missing orographic gravity wave drag near 60°S the cause of the stratospheric zonal wind biases in chemistry–climate models?

Is Missing Orographic Gravity Wave Drag near 60°S the Cause of the Stratospheric Zonal Wind Biases in Chemistry–Climate Models?

CHARLES MCLANDRESS AND THEODORE G. SHEPHERD

Department of Physics, University of Toronto, Toronto, Ontario, Canada

SAROJA POLAVARAPU

Climate Research Division, Environment Canada, Toronto, Ontario, Canada

STEPHEN R. BEAGLEY

Department of Earth and Space Science and Engineering, York University, Toronto, Ontario, Canada

(Manuscript received 10 June 2011, in final form 3 October 2011)

ABSTRACT

Nearly all chemistry–climate models (CCMs) have a systematic bias of a delayed springtime breakdown of the Southern Hemisphere (SH) stratospheric polar vortex, implying insufficient stratospheric wave drag. In this study the Canadian Middle Atmosphere Model (CMAM) and the CMAM Data Assimilation System (CMAM-DAS) are used to investigate the cause of this bias. Zonal wind analysis increments from CMAM-DAS reveal systematic negative values in the stratosphere near 60°S in winter and early spring. These are interpreted as indicating a bias in the model physics, namely, missing gravity wave drag (GWD). The negative analysis increments remain at a nearly constant height during winter and descend as the vortex weakens, much like orographic GWD. This region is also where current orographic GWD parameterizations have a gap in wave drag, which is suggested to be unrealistic because of missing effects in those parameterizations. These findings motivate a pair of free-running CMAM simulations to assess the impact of extra orographic GWD at 60°S. The control simulation exhibits the cold-pole bias and delayed vortex breakdown seen in the CCMs. In the simulation with extra GWD, the cold-pole bias is significantly reduced and the vortex breaks down earlier. Changes in resolved wave drag in the stratosphere also occur in response to the extra GWD, which reduce stratospheric SH polar-cap temperature biases in late spring and early summer. Reducing the dynamical biases, however, results in degraded Antarctic column ozone. This suggests that CCMs that obtain realistic column ozone in the presence of an overly strong and persistent vortex may be doing so through compensating errors.

1. Introduction

A too-cold Southern Hemisphere (SH) wintertime stratospheric polar vortex that breaks down too late in the year is a bias common to nearly all chemistry–climate models (CCMs). This so-called cold-pole bias is a long-standing problem, and despite significant improvements in models it continues to remain a serious impediment to progress (Erying et al. 2010, chapter 4; Butchart et al. 2011). These wind and temperature biases have important implications. First, they result in simulated ozone trends in the Antarctic middle stratosphere that are offset by several weeks with respect to observations (Stolarski et al. 2006). Second, since Antarctic ozone depletion is the primary driver of recent SH summertime climate change (e.g., Arblaster and Meehl 2006; McLandress et al. 2011), a delay in the breakdown of the Antarctic vortex will have an impact on the timing of the simulated tropospheric response. These reasons underscore the importance of alleviating these model biases, particularly if accurate predictions of future Antarctic lower-stratospheric ozone and SH climate change are to be obtained.

The fact that nearly all CCMs suffer from a cold-pole bias in SH winter and spring points to missing

Corresponding author address: Charles McLandress, Department of Physics, University of Toronto, 60 St. George St., Toronto ON M5S 1A7, Canada.
E-mail: charles@atmosp.physics.utoronto.ca

DOI: 10.1175/JAS-D-11-0159.1

stratospheric wave drag. This could be small-scale gravity wave drag (GWD), resolved (planetary) wave drag, or a combination of the two. Since both forms of wave drag respond to the mean winds, there is a strong coupling between them (McLandress and McFarlane 1993) and it is not possible to identify the root cause of the bias from the climate simulations alone. On the other hand, the 100-hPa meridional eddy heat flux (the typically used metric for tropospheric wave forcing of the stratosphere) in the CCMs seems, if anything, too strong at SH high latitudes in midwinter (see Erying et al. 2010, chapter 4), suggesting that the resolved waves are unlikely to be the cause of the bias. The general consensus therefore is that there is insufficient stratospheric GWD in the SH winter and spring. Although model biases in radiative heating could in principle be responsible, this seems unlikely since no common model bias was found in the radiation codes used in current CCMs (see Erying et al. 2010, chapter 3). Moreover, radiation codes can be tested against the ''truth,'' as determined by a line-by-line code. Since there is no such truth for GWD parameterizations, they are by far the most likely culprit.

So what is it in the current orographic and nonorographic GWD parameterizations that could cause the cold-pole bias in the SH stratosphere? To help answer that question, one must consider the two schemes separately. Although it is generally assumed that (parameterized) orographic GWD is far more important in the Northern Hemisphere because of the far greater amount of mountainous terrain, it does have an important impact on the SH, as will be shown later in our simulations. One important difference between the two hemispheres is the absence of mountains around the 60°S latitude circle, with the exception of several small isolated islands that are not resolved by CCMs. This results in a gap in parameterized orographic GWD in the stratosphere near 60°S, which is flanked in winter and spring by regions of strong orographic GWD arising from gravity waves generated by the Antarctic Peninsula to the south and by the Andes to the north.

Concerning the nonorographic GWD parameterizations, which treat small-scale gravity waves generated by sources other than mountains, there is considerable variation from model to model in the types of schemes and the gravity wave source spectra in the troposphere. Because of a dearth of observations needed to constrain the nonorographic source spectra, modelers tend to choose values that give reasonable mesospheric zonal mean winds and temperatures, with less regard for the stratosphere. Although mesospheric GWD can warm the polar winter stratosphere as a result of ''downward control'' and the long radiative time scales there (Garcia and Boville 1994), large temperature biases still persist even after the inclusion of nonorographic GWD schemes (Austin et al. 2003; Eyring et al. 2006). Moreover, the wide variety of parameterizations in use in CCMs makes it unlikely that nonorographic GWD could be the source of the common model bias.

Current GWD parameterizations, orographic and nonorographic alike, use a number of common simplifying assumptions to make them computationally efficient. One of the most important assumptions is that the waves propagate straight up. However, high-resolution model simulations of the SH polar vortex indicate that resolved small-scale gravity waves also propagate meridionally and are focused into the jet core as they propagate up into the stratosphere (Sato et al. 2009). The Sato et al. study reveals that mountain waves forced over the Antarctic Peninsula and southern Andes are focused toward 60°S, which suggests that the gap near 60°S in parameterized orographic GWD in the winter and spring stratosphere is unphysical. Another possible source of missing GWD is from small isolated mountainous islands in the Southern Ocean near 60°S, which can produce large momentum fluxes, which in turn would imply large local drag forces in the stratosphere (Alexander et al. 2009). Current orographic GWD parameterizations do not account for these islands because the grid cells containing them are treated as ocean. In both cases, neglect of these effects is common to all CCMs and thus represents a potential common source of systematic error.

In this paper we address the question of whether missing orographic GWD in the stratosphere around 60°S might account for the SH stratospheric zonal wind biases in CCMs, especially the delayed breakdown of the polar vortex. We do this through a two-step process, using the Canadian Middle Atmosphere Model (CMAM). The first step is to determine whether the model bias still exists when the resolved waves are represented correctly. We do this by constraining CMAM to follow observations through data assimilation and looking at the errors in the short-term forecasts. The assumption is that any systematic error in the short-term forecasts is attributable to ''fast'' physics in the model, which in the stratosphere is the parameterized GWD. We will show that the zonal wind analysis increments (which act to correct the forecast biases with respect to observations; see section 2b for the exact definition) in the CMAM Data Assimilation System (CMAM-DAS) exhibit a large negative maximum in the upper stratosphere near 60°S in the winter and early spring, which descends in time as the polar vortex weakens and begins to break up. This strongly suggests that missing GWD, in particular drag from gravity waves with near-zero ground-based phase speeds, is the cause of the delayed breakdown of the SH stratospheric polar vortex in CCMs.

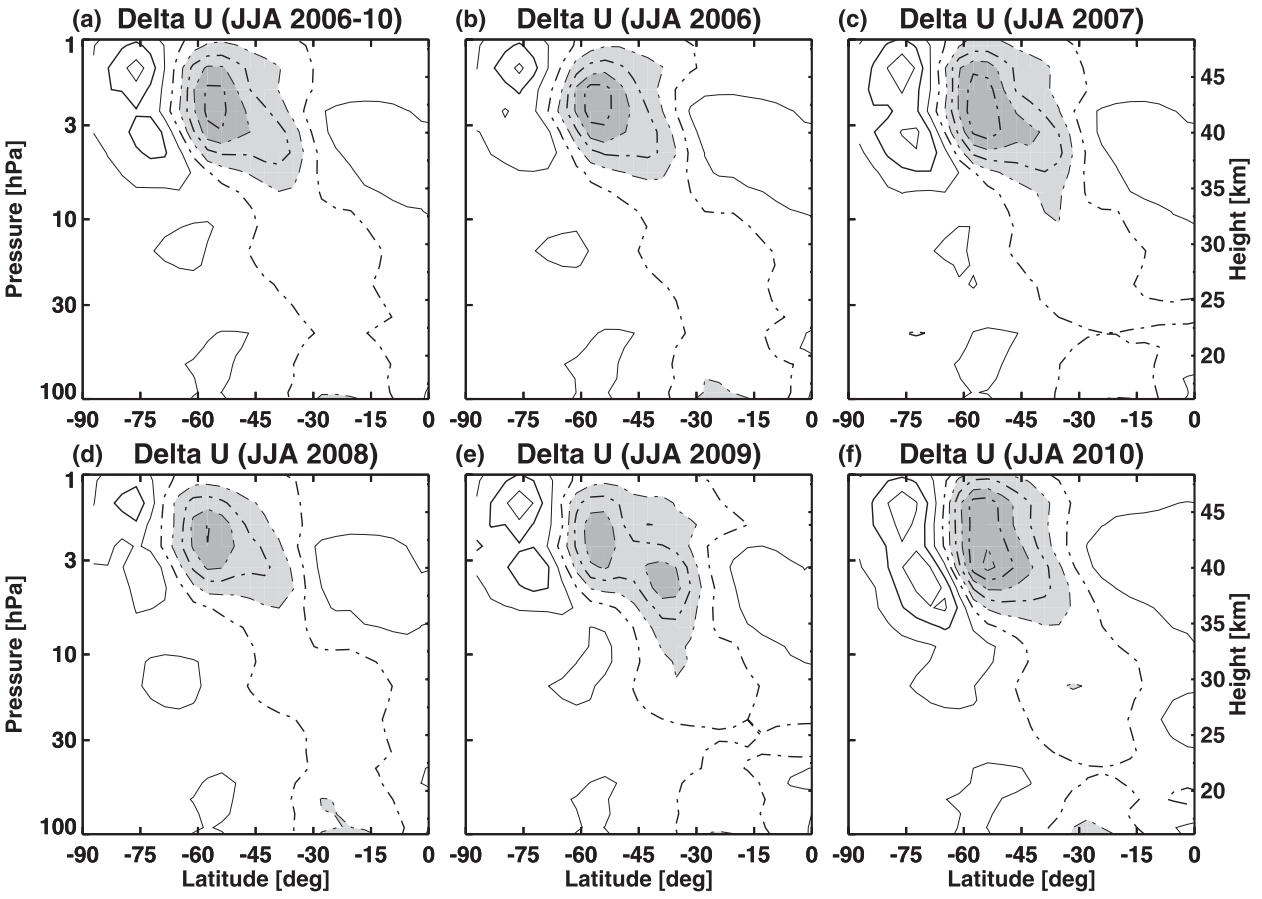

FIG. 1. Zonal mean zonal wind analysis increments from CMAM-DAS for JJA: (a) 5-yr average (2006–10) and (b)–(f) individual years. The increments have been expressed as a tendency by multiplying them by 4 day^{-1}. Contour interval is 0.6 m s^{-1} day^{-1}; negative values are dashed. Increments less than -0.9 and -2.1 m s^{-1} day^{-1} are shaded light and dark gray, respectively. The right axis shows log-pressure height, computed using a 7-km scale height.

The second step, which is motivated by the above findings and by the fact that the total (parameterized) GWD in CMAM-DAS is very weak near 60°S, is to perform a pair of free-running climate simulations using CMAM: a control experiment having the exact same configuration as the forecast model used in CMAM-DAS and an experiment with extra GWD applied at 60°S. Since the evidence points to orographic GWD as being the most likely source of the missing GWD, an orographic parameterization is used to apply the extra drag. We will show that the inclusion of this extra GWD significantly improves the simulation, resulting in more realistic zonal mean winds and temperatures in the SH stratosphere in winter and spring and a vortex that breaks down closer to the observed date. In response to the extra GWD there are also changes in resolved wave drag, which act to spread out the total wave drag (i.e., resolved plus total GWD) in latitude. Moreover, the earlier breakdown of the vortex reduces the resolved wave drag in the stratosphere in late spring and early summer, which reduces the stratospheric temperature biases in those months.

The paper is organized as follows. Section 2 gives a brief description of CMAM and CMAM-DAS. Section 3 discusses the use of the zonal wind analysis increments to infer missing GWD and presents results from CMAM-DAS for 2006–10. Section 4 discusses the free-running CMAM simulations. Section 5 summarizes our findings and discusses some implications of our results.

2. Models

Our study makes use of the Canadian Middle Atmosphere Model and the data assimilation system that is based on it. Here, a brief description of the two is given.

a. Canadian Middle Atmosphere Model

CMAM is a chemistry–climate model that extends from the earth's surface to about 100 km. The version used here employs a triangular spectral truncation of T47

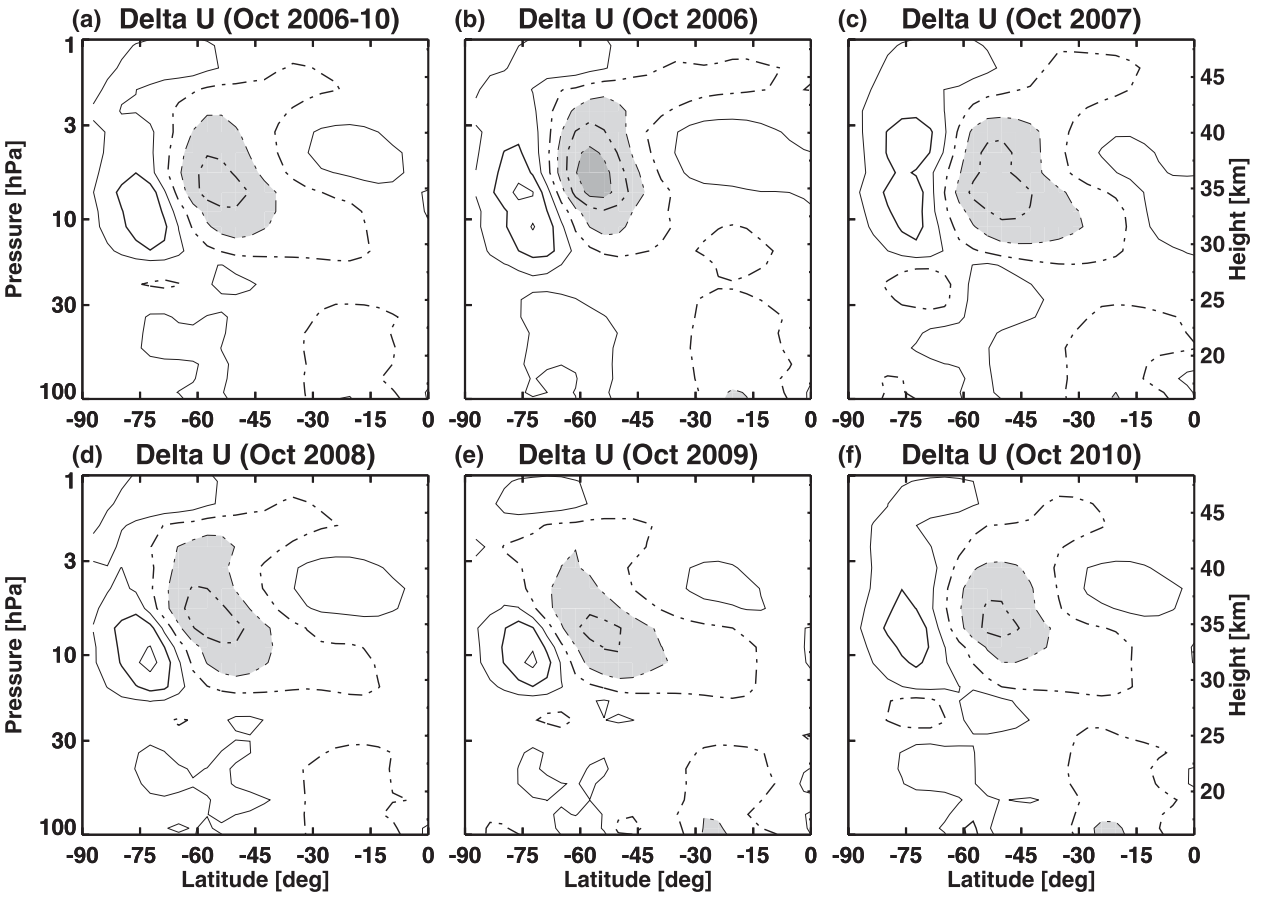

FIG. 2. As in Fig. 1, but for October.

in the horizontal, which corresponds to a 3.75° grid on which the "physics" calculations are performed. There are 71 levels in the vertical, with a resolution ranging from several tens of meters in the lower troposphere to about 2.5 km in the mesosphere. The stratospheric chemistry scheme is the one used for phase 1 of the CCM Validation Activity (CCMVal-1; Eyring et al. 2006), and this version of CMAM also includes simplified tropospheric chemistry [see Ren et al. (2011) for more details]. Detailed descriptions of the model parameterizations and the stratospheric chemistry scheme are provided in Scinocca et al. (2008) and de Grandpré et al. (2000), respectively.

To provide the forecasts required to generate the CMAM-DAS analyses, which span the period from October 2005 to December 2010, the chemical tracers are "spun up" using a 10-yr time-slice simulation for 1990 conditions. The model is then integrated from 1990 to 2005 using transiently evolving observed sea surface temperatures (SSTs) and chemical boundary conditions from CCMVal-1 (Eyring et al. 2005). The resulting chemical and dynamical fields from October 2005 are used to initialize the very first forecast in the CMAM-DAS assimilation. The setup of the two free-running CMAM simulations that are performed to assess the impact of extra GWD at 60°S is discussed in section 4a.

b. CMAM Data Assimilation System

CMAM-DAS is a three-dimensional variational data assimilation (3DVar) system, which uses CMAM as the underlying forecast model. An earlier version of the system is described in Polavarapu et al. (2005a). Two significant improvements to the system have since been made. The digital filter has been replaced by an incremental analysis updating scheme (Polavarapu et al. 2004), and observations are now compared to background (i.e., forecast) fields at the closest hour during the 6-h assimilation cycle using a first guess at appropriate time (FGAT) scheme (see Ren et al. 2011). Standard meteorological measurements are assimilated up to 1 hPa. Above 10 hPa only the channel 10–13 radiances from the Advanced Microwave Sounding Unit-A (AMSU-A) from the National Oceanic and Atmospheric Administration (NOAA) satellites *NOAA-15* and

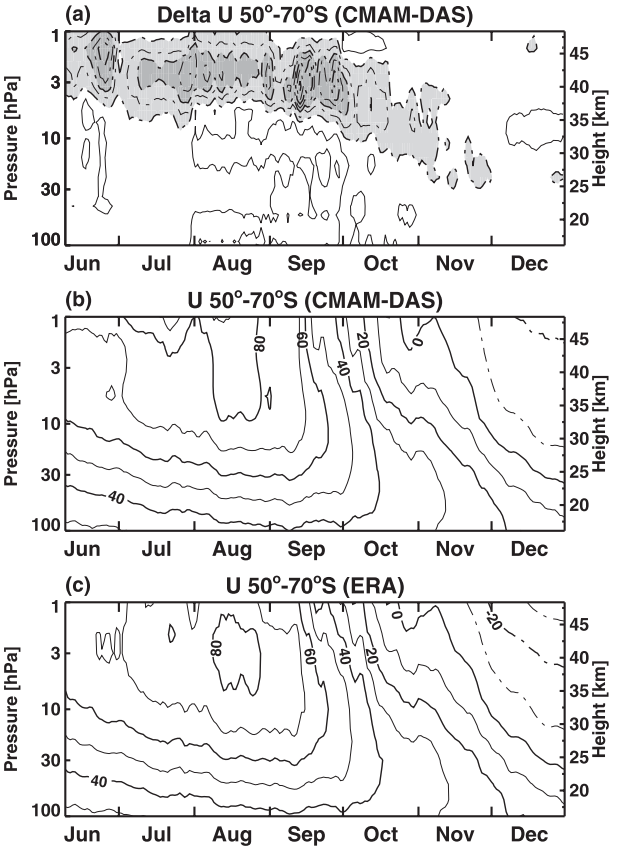

FIG. 3. Daily (a) zonal wind analysis increments from CMAM-DAS, (b) CMAM-DAS zonal wind, and (c) ERA-Interim zonal wind, all averaged from 50° to 70°S for 2006–10. The increments have been smoothed using a 3-day running mean and expressed as a tendency by multiplying them by 4 day^{-1}. Contour intervals are 0.8 m s^{-1} day^{-1} in (a) and 10 m s^{-1} in (b) and (c); negative values are dashed. Light and dark shadings in (a) are used for values less than -0.4 and -2.0 m s^{-1} day^{-1}, respectively.

NOAA-16 are assimilated. Although ozone and other trace gases are predicted by the model, they are not assimilated.

An assimilation cycle centered at time t consists of the following steps. First, a 6-h forecast that is initialized from the previous analysis cycle, which is centered at $t - 6$ h, is performed from $t - 3$ to $t + 3$ h. Differences between the forecast (saved at 1-h intervals) and the observations are then computed in observation space at the appropriate times in that 6-h window. The differences in observation space are transformed to model space using the error covariances to yield differences in the prognostic variables valid at time t. These latter differences, which we will refer to as ''analysis increments,'' are divided by $N\Delta t$, where N is the number of model time steps in the 6-h window and Δt is the model time step, and used as forcing terms for the model when it is rerun from $t - 3$ to $t + 3$ h. This second integration yields the analysis. This process is repeated using the analysis at $t + 3$ h as the initial conditions for the next cycle. We use CMAM-DAS analyses for 2006–10.

3. Inferring missing GWD from analysis increments

Because the assimilation scheme assumes no biases, persistent analysis increments indicate the presence of biases in the forecast model, the observations, or both (Daley 1991). To isolate these biases, the increments must be averaged in both space and time. This removes the random component due to nonlinear error growth (i.e., the ''butterfly effect''), as well as the random errors in the observations and the effect of the irregular spatial sampling pattern of the observations.

Since the model biases arise over 6 h, they must be due to fast processes, which in the stratosphere means parameterized GWD. Although it is possible that the bias reflected in the analysis increments is due to the AMSU-A observations, which are the only observations at these heights, we hypothesize that the bias is model related because its latitudinal and temporal evolution is related to the vortex evolution. We will then verify a posteriori that this assumption is justified.

Figures 1 and 2 show zonal mean zonal wind analysis increments from CMAM-DAS for June–August (JJA) and October, respectively. The wind increments have been multiplied by 4 day^{-1} (i.e., one analysis increment field every 6 h) in order to express them as a tendency for later comparison to the parameterized GWD tendencies. Note the region of large negative increments centered at about 55°S and 3 hPa in JJA, which descends to about 5 hPa in October. This feature is robust since it occurs each year, at the same place, and with approximately the same strength. Negative increments indicate that the zonal mean zonal wind tendencies in the forecast model (i.e., CMAM) are too strong, meaning that the insertion of the observations is acting to decelerate those winds. The tapering off of the analysis increments above about 2 hPa in Fig. 1 occurs because the background-error covariance matrix, which spreads information both horizontally and vertically (see Daley 1991), is reduced above 2 hPa to prevent spurious analysis increments from appearing far above the region of observations, which cannot be justified statistically (see section 3a of Polavarapu et al. 2005b).

Centered near 75°S in the stratosphere in Figs. 1 and 2 is a region of weak positive wind increments, which could be an indication of too much GWD at that location. They could also reflect the near-instantaneous (i.e., within a 6-h period) nonlocal response to the

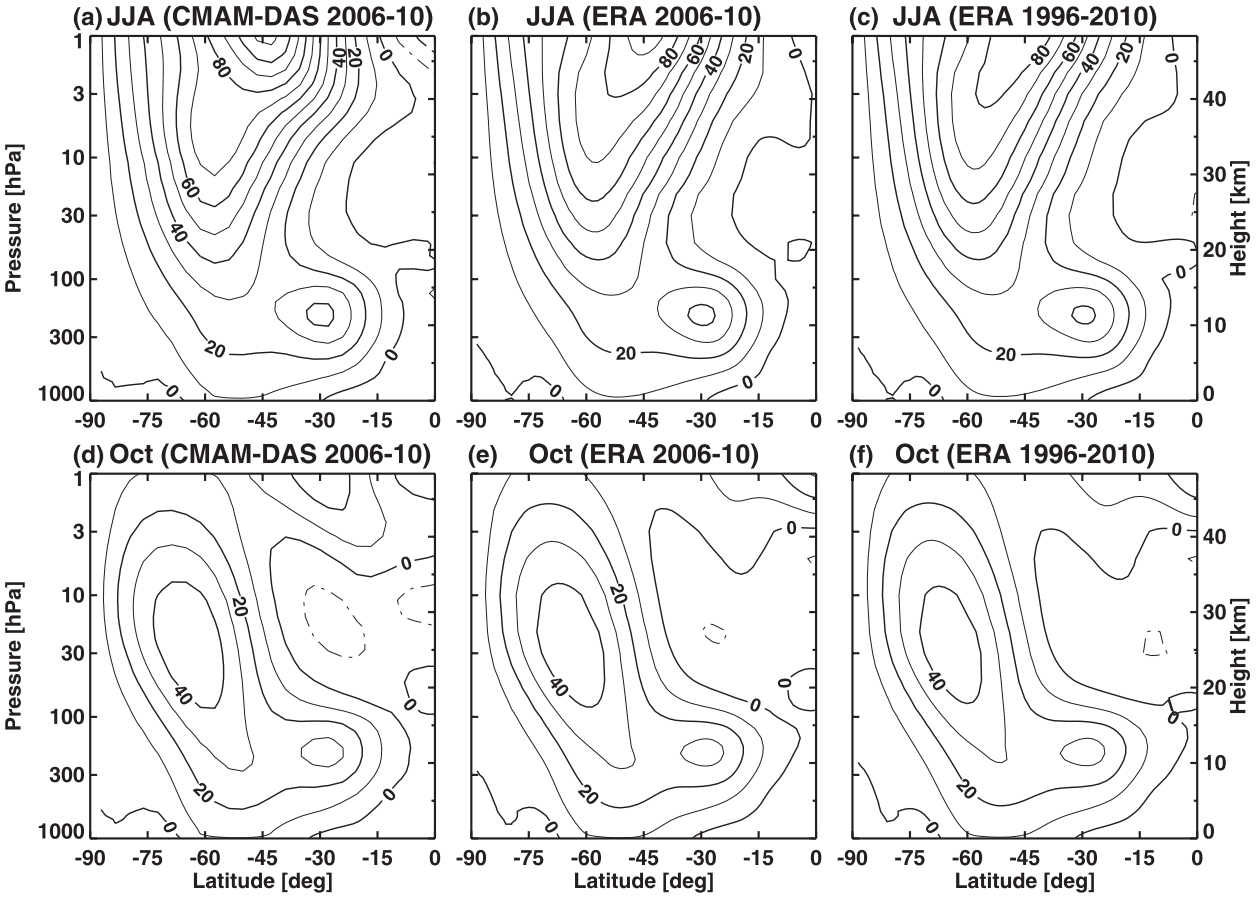

FIG. 4. Zonal mean zonal wind for (a)–(c) JJA and (d)–(f) October, showing results for (a),(d) CMAM-DAS for 2006–10, and ERA-Interim for (b),(e) 2006–10 and (c),(f) 1996–2010. Contour interval is 10 m s^{-1}.

CMAM-DAS increments at 60°S, which the model would try to spread out, causing such a dipole structure. Whatever the cause of the positive increments at high latitudes, we will ignore them since they are substantially weaker than the negative increments near 60°S.

Figure 3a shows 5-yr average daily time series of the zonal wind analysis increments averaged from 50° to 70°S. The increments have been smoothed using a 3-day running mean to help eliminate the short time scale variations, which are not part of the systematic bias. Over the course of the winter the height of the (negative) increment maximum remains nearly constant. During spring, the maximum descends in time, disappearing in November. Since the weighting functions for the different AMSU-A channels peak at different heights, it seems unlikely that a bias in the AMSU-A data could produce such a seasonally evolving height-dependent structure in the increments. Figure 3b shows the corresponding zonal mean zonal winds from CMAM-DAS. The descent of the increment maximum in October and November when the winds weaken and the vertical shear of the winds becomes negative, as well as the absence of negative increments above the zero-wind line, are consistent with GWD produced by waves having near-zero ground-based phase speeds, such as orographic gravity waves.

Figure 3c shows zonal mean zonal winds from the interim version of the European Centre for Medium-Range Weather Forecasting (ECMWF) Re-Analysis (ERA-Interim; Dee et al. 2011) for the years 2006–10 and averaged from 50° to 70°S. Below about 3 hPa the agreement with the CMAM-DAS winds (Fig. 3b) is excellent. Above, the differences are larger, but the two analyses are still in good agreement. The two left-hand columns in Fig. 4 show the zonal mean zonal winds for JJA and October for CMAM-DAS and ERA-Interim for 2006–10. With the exception of the region above 3 hPa and equatorward of 60°S, the agreement between the two is excellent. The differences in the upper levels are due to differences in the data assimilation systems and the underlying forecast models. However, the zonal wind differences are largest where the CMAM-DAS

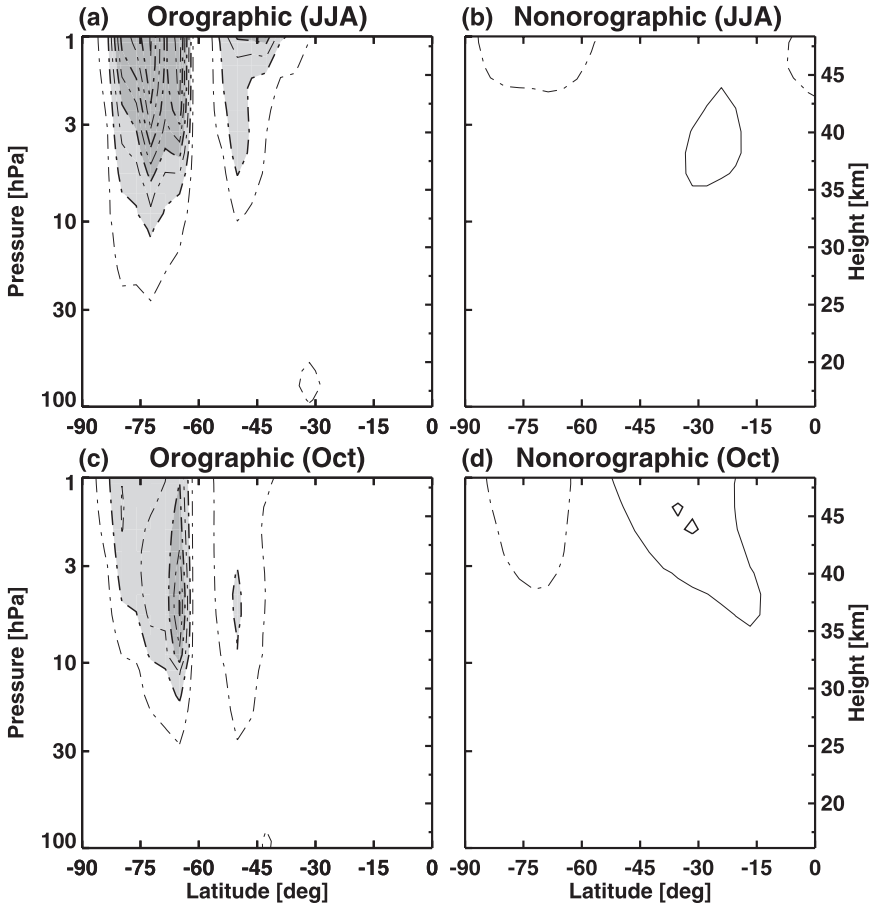

FIG. 5. Zonal mean zonal component of (a),(c) orographic and (b),(d) nonorographic GWD for (a),(b) JJA and (c),(d) October for 5-yr average CMAM-DAS. Contour interval is 0.6 m s^{-1} day^{-1}; negative values are dashed. Light and dark shadings are used for values less than -0.9 and -2.1 m s^{-1} day^{-1}, respectively.

zonal wind analysis increments are small; in the region near 60°S where the large negative increments occur, the zonal wind differences are small. In October the CMAM-DAS increments peak lower down, around 5 hPa (Fig. 2). Yet at these altitudes the CMAM-DAS winds agree very well with ERA-Interim.

Although we cannot rule out the possibility that the presence of the systematic analysis increments is due to a bias in the observations, it seems unlikely that all the various AMSU-A platforms available during 2006–10 would have the same bias. Indeed, operational centers apply separate bias corrections to each instrument on each platform. Since ECMWF assimilated all of the available AMSU-A datasets (see Fig. 14 of Dee et al. 2011), whereas CMAM-DAS assimilated only two, the agreement with the ERA-Interim winds (Fig. 3) at the height of the peak increments suggests that observation bias is not the primary cause of the systematic zonal wind analysis increments in CMAM-DAS.

The right column in Fig. 4 shows zonal mean zonal winds for ERA-Interim averaged from 1996–2010. This 15-yr period is chosen because it is long enough to yield a reasonable climatology yet avoids the years when the ozone hole was deepening, during which there was a trend in the breakdown date of the SH vortex (e.g., Waugh et al. 1999). We have verified that there is no detectable trend in the final warming dates (as defined in the next section) at 50 and 10 hPa. The close agreement between the 15- and 5-yr averages (Figs. 4b,e) indicates that the latter are representative of the longer-term climatology, which suggests that the CMAM-DAS zonal wind analysis increments are also representative of a longer-term climatology of the model bias (if it were available). In the next section we use ERA-Interim data for comparison purposes since they span a longer time period than CMAM-DAS, thus enabling more accurate climatological biases in the free-running CMAM simulations to be obtained.

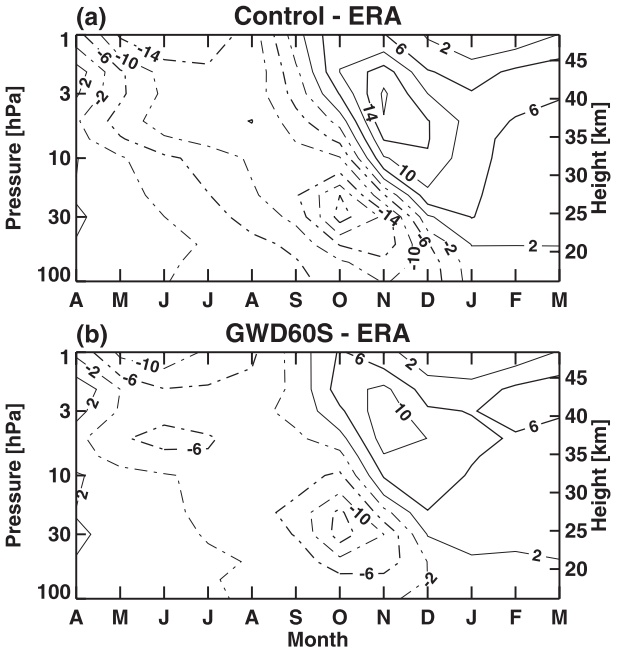

FIG. 6. Monthly mean temperature biases with respect to ERA-Interim, averaged from 70° to 90°S: (a) control experiment and (b) GWD60S experiment. ERA-Interim data are for 1996–2010. Contour interval is 4 K.

4. Free-running CMAM simulations

The above findings support the hypothesis that CMAM underestimates the amount of GWD in the stratosphere near 60°S during winter and early spring. In this section we describe a pair of free-running climate simulations using CMAM, one with extra GWD applied at 60°S and the other without. To gain insight into how to include this extra drag, it is helpful to first examine the parameterized GWD in CMAM-DAS. Figure 5 shows the orographic and nonorographic GWD for JJA and October, which are parameterized using the schemes of Scinocca and McFarlane (2000) and Scinocca (2003), respectively. Comparing the two, one immediately sees that the orographic GWD far outweighs the nonorographic GWD in the region of interest (i.e., mid- to high latitudes). Moreover, the gap in orographic GWD near 60°S occurs close to where the analysis increments in Figs. 1 and 2 are most negative. As argued in the introduction, the likelihood that the missing wave drag in that region is orographic GWD suggests that a modified version of an existing orographic GWD parameterization would be a reasonable way of including the extra drag. Note also that the negative zonal wind analysis increments in the stratosphere near 60°S are nearly as large as the parameterized orographic GWD in neighboring latitudes, further indication that the increments reflect missing GWD.

a. Experimental setup

The first simulation is the control experiment, which is identical in configuration to the version of CMAM used as the forecast model in CMAM-DAS. To perform this simulation using the same SSTs as in CMAM-DAS, only the period from January 2006 to February 2011 is considered. To obtain a reasonably accurate climatology of the breakdown of the SH vortex, a set of 12 ensemble members per experiment is performed, each spun off from slightly different initial conditions. This yields a total of 60 yr of simulations.

The second simulation is the GWD60S experiment, which also comprises an ensemble of twelve 5-yr simulations. It is identical to the control experiment, but with extra GWD applied at 60°S using a modified version of the orographic GWD parameterization of McFarlane (1987). This is achieved using a horizontally uniform subgrid-scale topographic height standard deviation field of 1000 m and a latitude-dependent momentum flux factor [Eq. (3.1b) of McFarlane (1987)] having a Gaussian profile centered at 60°S with a maximum value of 10^{-7} m^{-1} and an e-folding width of 5°. The extra drag is applied only to the zonal component of the flow. These settings were chosen to yield a zonally averaged GWD of about -2 m s^{-1} day^{-1} at 60°S near 3 hPa in winter, which is the value of the missing drag inferred from CMAM-DAS (Figs. 1 and 3a). The momentum flux at the source level, which is located one model level above the earth's surface, is computed from Eq. (2.30) of McFarlane (1987) and is estimated to be about 10 mPa in the zonal mean using values of 10 m s^{-1}, 0.01 s^{-1}, and 1 kg m^{-3} for surface wind, buoyancy frequency, and density, respectively. It must be emphasized that the particular choice of parameter settings used in the GWD60S experiment is not intended to be a realistic representation of gravity wave sources, but only a means of adding, in a simple yet physically consistent manner, the missing GWD inferred from CMAM-DAS. Nevertheless, the above zonal mean surface momentum flux is consistent with the observed flux of 200 mPa from South Georgia Island averaged over a 2.75° horizontal grid box (Alexander et al. 2009) if several such isolated mountainous islands were to generate such a flux.

b. Control simulation

The top panel in Fig. 6 shows the monthly mean annual cycle of SH polar-cap temperature differences between the control simulation and the 15-yr ERA-Interim climatology. A cold bias of 2–15 K exists from June to October throughout the entire stratosphere. By the thermal wind relation, this results in an excessively strong polar jet, as seen in the left panels of Fig. 7, which show daily zonal

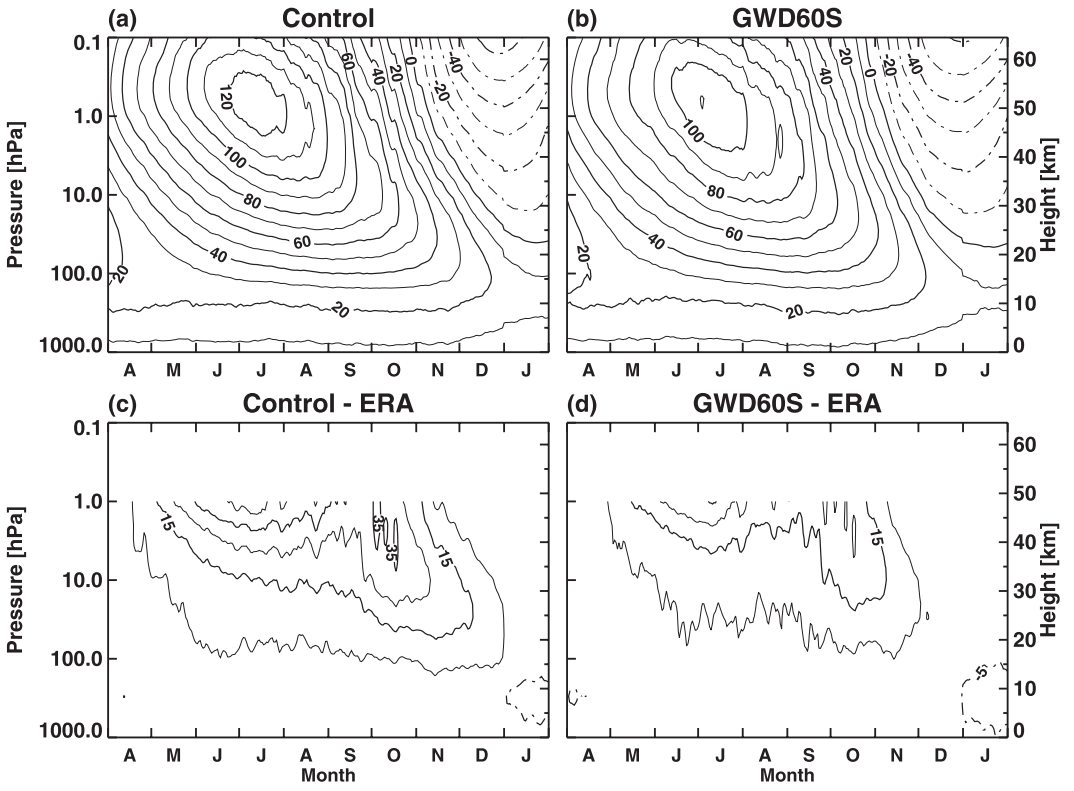

FIG. 7. Daily zonal mean zonal wind averaged from 50° to 70°S: (a) control experiment, (b) GWD60S experiment, (c) control minus ERA-Interim, and (d) GWD60S minus ERA-Interim. ERA-Interim data are for 1996–2010. Contour interval is 10 m s^{-1}.

winds from the control simulation averaged from 50° to 70°S and the corresponding biases with respect to the 15-yr ERA-Interim climatology. The winds in the control simulation exceed those in the reanalysis by up to 40 m s^{-1} and change to easterlies too late in the season, as seen by the positive biases in the lower stratosphere that persist at least through to the end of December.

The delayed breakdown of the SH vortex in the control simulation is better illustrated in Fig. 8, which shows the mean final warming dates at 50 hPa. These are computed using the method of Black and McDaniel (2007), which defines the final warming date as when the zonal mean zonal wind at 60°S and 50 hPa drops below 10 m s^{-1} and remains below that value through summer. The error bars denote the 95% confidence intervals of the means, computed using a standard *t* test. The mean final warming date for the control simulation is about 3 weeks later than that for ERA-Interim, with these differences being significant at the 95% level. This delay in the vortex breakdown is consistent with the CMAM results in Butchart et al. (2011), which is not surprising since the GWD schemes used in the two versions of the model are identical [although the version of CMAM used there had T32 resolution and included several important changes to the stratospheric chemistry scheme compared with the CCMVal-1 chemistry used here (Morgenstern et al. 2010)].

Figure 9 shows the vertical profile of the final warming dates, defined as when the zonal mean zonal wind at 60°S and a given pressure level drops below zero and remains easterly through summer. If this criterion is not met (e.g., some years in the lower stratosphere), a final warming date is not computed for that year and pressure level. The reason why this criterion is employed for this diagnostic, and not the one used in Fig. 8, is for consistency with previous results showing the descent of the zero-wind line (e.g., Eyring et al. 2006; Butchart et al. 2011). As in Fig. 8, the mean final warming dates for the control simulation (large red circles) are 15–20 days later than for ERA-Interim (large black circles), which is consistent with the results shown in Fig. 7. These differences in the mean warming dates are significant at the 95% level. Note the large amount of year-to-year variability, as seen by the spread in the small colored dots, which indicates that short simulations [e.g., 5 yr, as done in Scinocca et al. (2008)] are not sufficient to accurately compute the climatological final warming dates. Inferences about possible causes of the delayed

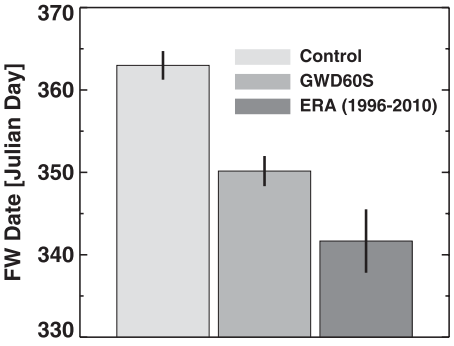

FIG. 8. Mean final warming dates at 50 hPa for the control experiment, GWD60S experiment, and ERA-Interim for 1996–2010. The error bars denote the 95% confidence intervals of the means. The final warming date is here defined by when the zonal mean zonal wind at 60°S and 50 hPa drops below 10 m s^{-1} and remains so for the duration of the summer.

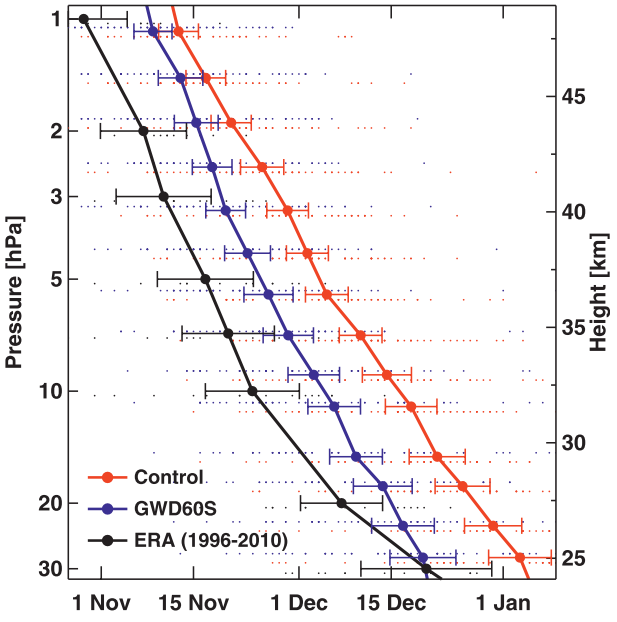

FIG. 9. Final warming dates vs pressure: control experiment (red), GWD60S experiment (blue), and ERA-Interim for 1996–2010 (black). The climatological means are given by the large circles (joined by lines) and individual years by small dots. The error bars denote the 95% confidence intervals of the means. The final warming date is here defined by when the zonal mean zonal wind at 60°S becomes easterly and remains so for the duration of the summer. The small dots have been offset slightly in the vertical for clarity.

breakdown of the SH polar vortex using short simulations should therefore be treated with some caution.

Before discussing the results of the GWD60S experiment, it is first necessary to demonstrate that the cold-pole bias in winter and the delayed vortex breakdown in spring in the control simulation are not due to a lack of resolved wave flux from the troposphere. For that, the 100-hPa meridional eddy heat flux (the typically used metric for tropospheric wave forcing) is used. Figure 10, which shows seasonally averaged results for midlatitudes for winter and spring, indicates that the heat flux for the control simulation is in good agreement with the ERA-Interim results.

c. *Impact of extra GWD*

Inclusion of the extra GWD at 60°S results in a substantial reduction in the wind and temperature biases as seen in Figs. 6 and 7. The temperature biases in the GWD60S experiment are much smaller than in the control simulation, with improvement even occurring in early summer in the stratosphere when the direct impact of the extra GWD is negligible (Fig. 6b). Moreover, the date of the transition to easterlies also has improved, as seen by the lack of contours in the stratosphere in December in Fig. 7d. The latter is quantified in Figs. 8 and 9, which

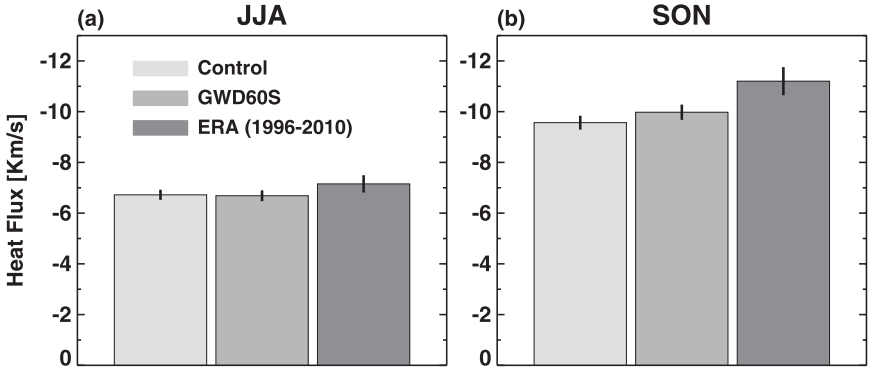

FIG. 10. Meridional heat flux at 100 hPa averaged from 40° to 80°S for (a) JJA and (b) September–November (SON). From left to right in each panel are shown the control experiment, GWD60S experiment, and ERA-Interim for 1996–2010. The error bars denote the 95% confidence intervals of the means.

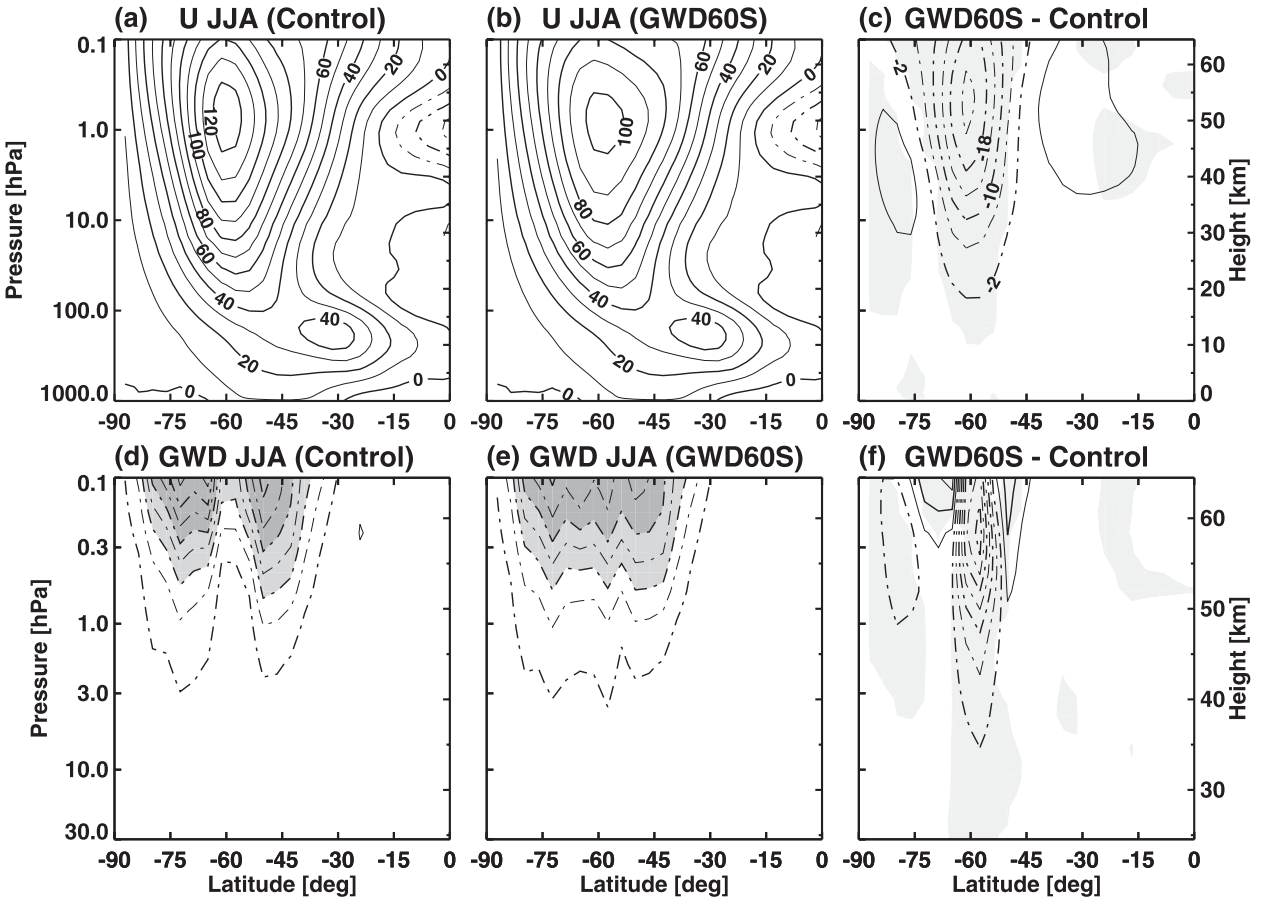

FIG. 11. (a)–(c) Zonal mean zonal wind and (d)–(f) total GWD for JJA: (a),(d) control experiment, (b),(e) GWD60S experiment, and (c),(f) their differences (GWD60S minus control). Contour intervals are 10 m s^{-1} in (a) and (b), 4 m s^{-1} in (c), 4 m s^{-1} day^{-1} in (d) and (e), and 2 m s^{-1} day^{-1} ($\pm 1, 3, \ldots$) in (f). Light and dark shadings in (d) and (e) are used for values less than -10 and -18 m s^{-1} day^{-1}, respectively. Shading in (c) and (f) denotes regions where the differences are statistically significant at the 95% level. Note that different vertical ranges are used in the top and bottom panels. Total GWD is the sum of the orographic and nonorographic GWD, and includes the extra GWD in (e).

show that the final warming dates in the GWD60S experiment are about 2–3 weeks earlier than in the control experiment and are now in closer agreement with ERA-Interim. Although only this one experiment was performed using the extra GWD at 60°S, it is clear that if different parameter settings had been used that either increased the GWD at this latitude or broadened its latitudinal extent, then the winter polar vortex would warm more and break down earlier than in the GWD60S experiment.

The latitude–height structure of the zonal mean zonal wind and total GWD (i.e., sum of orographic, nonorographic, and extra GWD) for JJA and October are shown in Figs. 11 and 12, respectively. The shading in the difference plots in these and subsequent figures denotes the 95% confidence levels, computed using a standard *t* test. As discussed above, the inclusion of the extra GWD results in weaker winds in the stratosphere at 60°S.

d. Resolved wave drag feedbacks

The inclusion of the extra GWD does not change significantly the resolved wave flux from the troposphere, as seen by the similarity of the 100-hPa meridional heat flux in the two experiments in Fig. 10. However, as a result of the changes in the zonal mean zonal winds induced by the extra GWD, changes in the resolved wave drag in the stratosphere do occur. This feedback is demonstrated in Fig. 13, which shows the resolved wave drag for JJA and October. Considering first the results for JJA, there is a tongue of strong wave drag at 60°S in the control experiment (Fig. 13a). This feature closely follows the jet axis (Fig. 11a), where the strong latitudinal curvature of the jet produces a strong potential vorticity gradient, which forms an efficient waveguide for the vertical propagation of planetary waves. In the GWD60S experiment (Fig. 13b), the tongue of wave

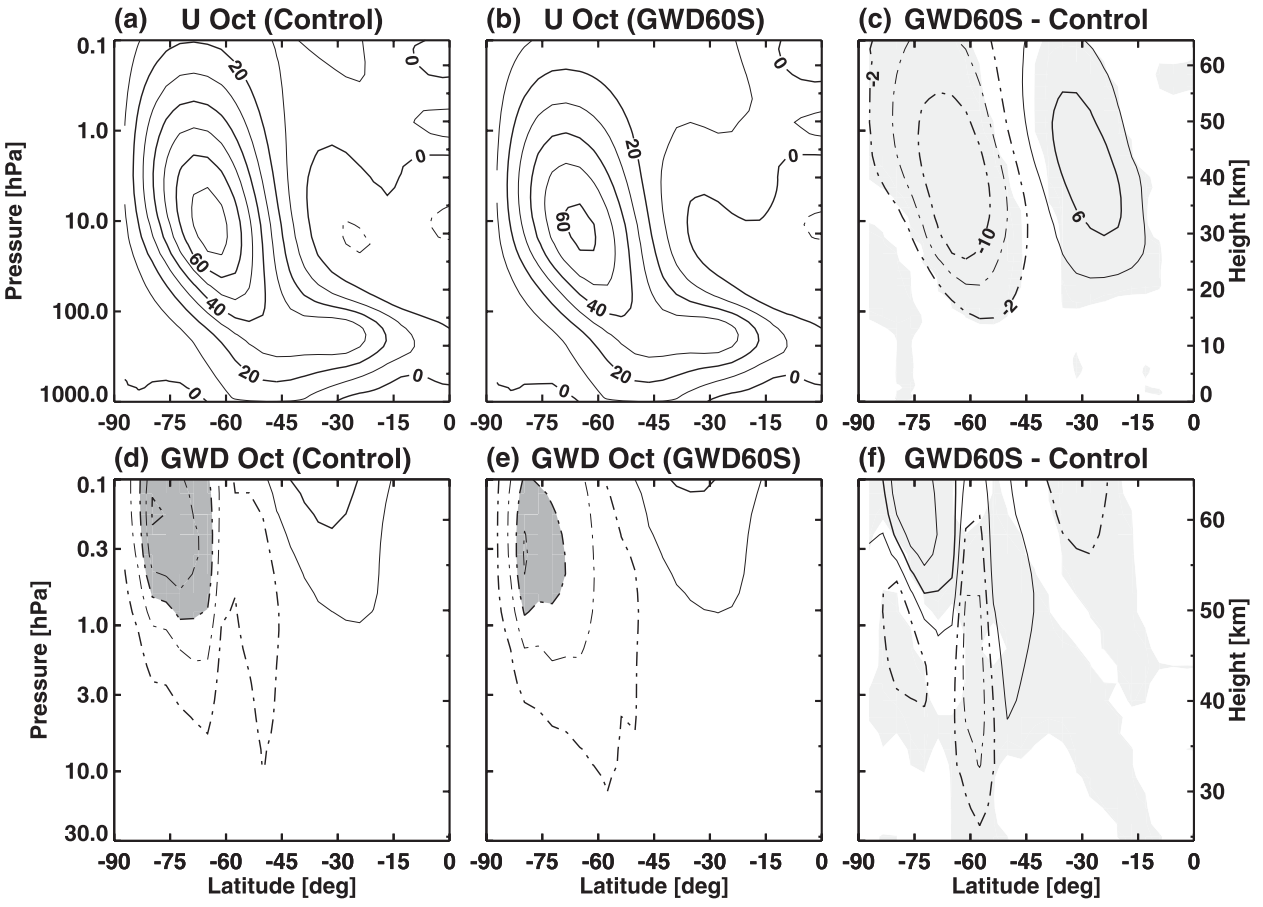

FIG. 12. As in Fig. 11, but for October. Contour intervals in (d)–(f) are half of those used in Figs. 11d–f. Shading in (d) and (e) is used for values less than -5 m s^{-1} day^{-1}.

drag at 60°S is absent and the resolved wave drag is spread out over a wider latitudinal region. This occurs because the zonal mean zonal winds are weaker at 60°S than in the control experiment, providing a less effective waveguide. This is demonstrated in the top panels of Fig. 14, which show the quasigeostrophic refractive index squared R^2 (Matsuno 1970) and the Eliassen–Palm (EP) flux for JJA. The corresponding differences in R^2 (GWD60S minus control; Fig. 14c) are negative along the jet axis near 60°S, indicating less favorable conditions for the vertical propagation of planetary waves in the GWD60S experiment, as is borne out by the downward pointing EP flux difference vectors near this latitude. A similar effect is also seen in October (bottom panels of Figs. 13 and 14), although it is weaker than in JJA because of the weaker winds.

The corresponding differences in the resolved wave drag between the two experiments (Figs. 13c,f) exhibit a meridional cell-like pattern, with positive values centered at 60°S and negative values to either side, which is indicative of a meridional spreading of the wave drag. This pattern is similar to that of the GWD differences (Figs. 11f and 12f) but is of opposite sign, and it results in some cancellation near 60°S when the differences in the total wave drag are computed, as seen in Fig. 15. Comparing this to the GWD differences (Figs. 11f and 12f), there is a weaker total wave drag difference near 60°S but larger values at adjacent latitudes, resulting in a latitudinal spreading of the wave drag difference.

In late spring and early summer when the extra GWD has little or no direct impact on the zonal mean zonal winds and temperatures [since the parameterized (stationary) orographic gravity waves cannot propagate past the zero-wind line], the above mechanism does not operate. Nevertheless, large changes in resolved wave drag are evident at this time. These changes are, in fact, responsible for the reduced polar-cap temperature biases in the stratosphere in late spring and early summer (Fig. 6). The left panels in Fig. 16 show the monthly mean annual cycle of the differences in polar-cap temperature and residual vertical velocity between the two

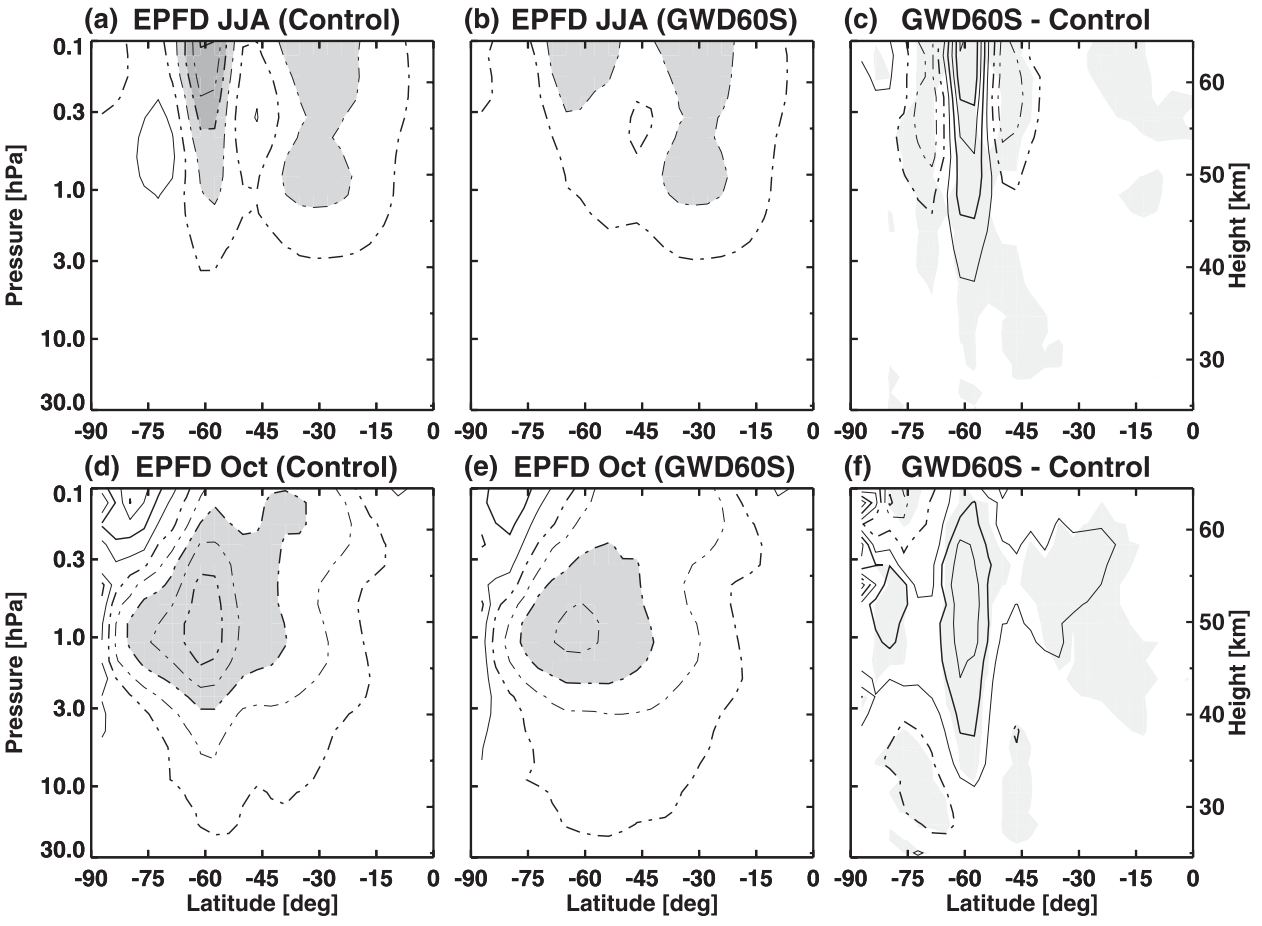

FIG. 13. Resolved wave drag for (a)–(c) JJA and (d)–(f) October: (a),(d) control experiment, (b),(e) GWD60S experiment, and (c),(f) their differences (GWD60S minus control). Contour intervals in the top and bottom rows are the same as in Figs. 11d–f and 12d–f, respectively. Light and dark shadings in (a) and (b) are used for values less than -6 and -10 m s^{-1} day^{-1}, respectively; and shadings in (d) and (e) for values less than -5 m s^{-1} day^{-1}. Shading in (c) and (f) denotes regions where the differences are statistically significant at the 95% level. Resolved wave drag is the Eliassen–Palm flux divergence expressed in units of force per unit mass.

experiments. The warming in winter/early spring and the cooling in late spring/early summer are a consequence of enhanced and reduced downwelling, respectively. The right panels in Fig. 16 show the change in total wave drag averaged over midlatitudes, along with the corresponding change in total GWD. From November to January the total wave drag changes are completely dominated by the resolved wave drag changes, the changes in total GWD being miniscule for the reason given above. The reduction in resolved wave drag drives the anomalous upwelling, which explains the anomalous cooling seen in Fig. 16a. The reduction in resolved wave drag results from the earlier breakdown of the vortex in the GWD60S experiment, which inhibits the vertical propagation of planetary waves into the stratosphere. This effect also explains the increase in the resolved wave drag at 50 hPa in November; that is, the lowering of the zero-wind line in the GWD60S experiment (Fig. 9) causes the planetary waves to dissipate lower down. These resolved wave drag changes and their impact on temperature are analogous (with a sign change) to the effect of the ozone hole (Manzini et al. 2003), which like the cold-pole bias acts to delay the breakdown of the vortex.

5. Summary and discussion

We have used spatially and temporally averaged zonal wind analysis increments from CMAM-DAS to infer missing stratospheric GWD in CMAM. The use of CMAM-DAS is crucial because it is only when one corrects the zonal winds and resolved wave drag (through data assimilation) that one can attribute the remaining bias to GWD. Persistent negative zonal wind analysis

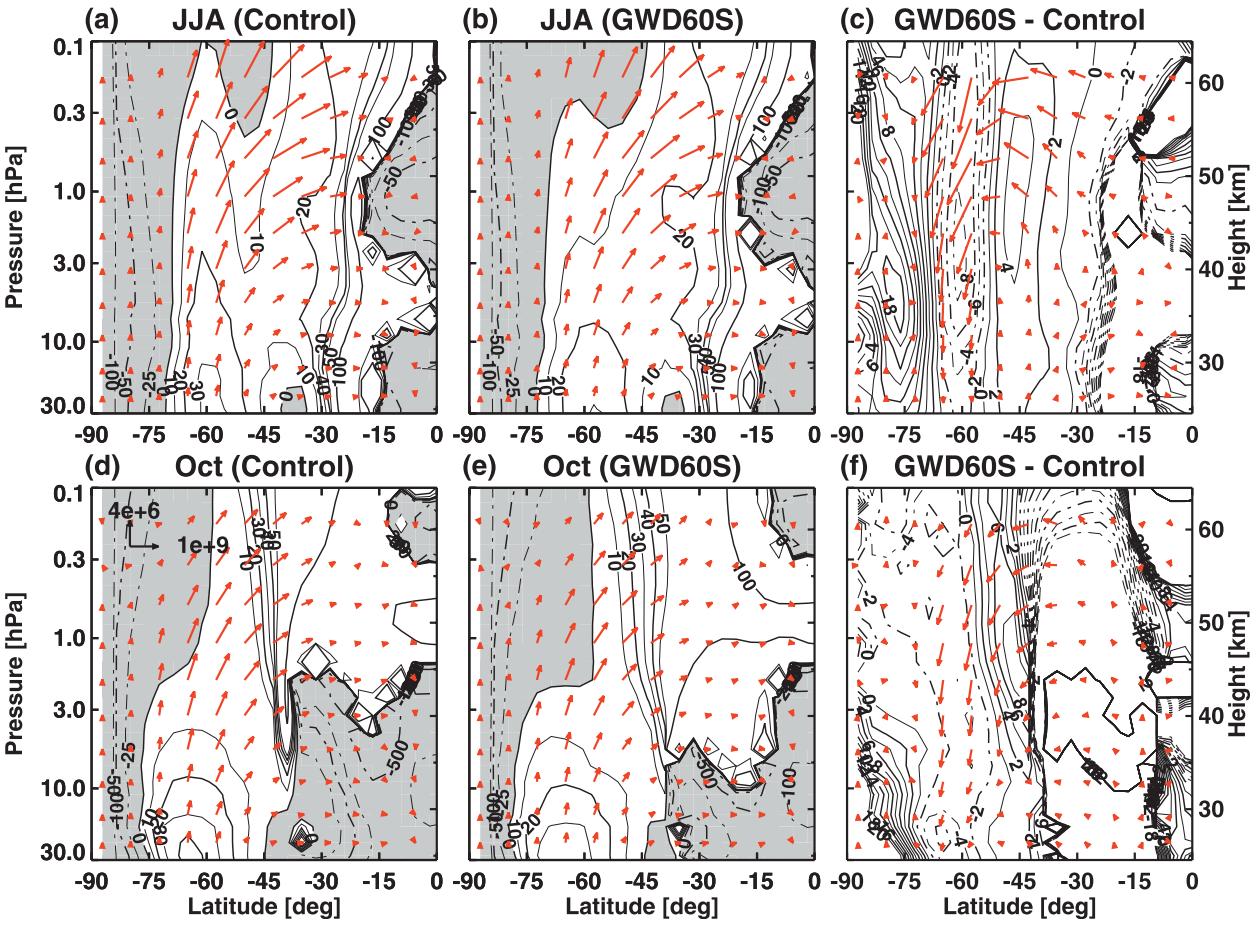

FIG. 14. Quasigeostrophic refractive index squared R^2 for stationary zonal wavenumber 1 and EP flux vectors for all zonal wavenumbers for (a)–(c) JJA and (d)–(f) October: (a),(d) control experiment, (b),(e) GWD60S experiment, and (c),(f) their differences (GWD60S minus control). The refractive index squared is computed using a buoyancy frequency of 0.0173 s^{-1} and a density scale height of 7 km, and is multiplied by the square of the earth's radius to make it nondimensional. In (a),(b),(d), and (e) negative values of R^2 are shaded; in (c) and (f) values of $|R^2| > 20$ are not plotted. The EP flux vectors are divided by the background density to highlight the upper stratosphere. The black arrows in the upper-left corner of (d) denote the magnitude of the EP flux components; the difference vectors in (c) and (f) are multiplied by a factor of 5 for visibility.

increments are found near 60°S in winter and early spring. The negative analysis increment maximum remains roughly constant in time over winter and descends as the zonal mean zonal winds weaken and the zero-wind line descends. This behavior is suggestive of orographic GWD, which is enhanced in regions where the background winds decrease with height, as discussed in McFarlane (1987).

The region of negative increments near 60°S happens to coincide with the region where the drag from current orographic GWD parameterizations is identically zero because only ocean grid cells are located in this latitude band. These two facts provide motivation for a pair of climate simulations using the free-running CMAM: a control simulation that is configured exactly as the forecast model in CMAM-DAS and a simulation with extra orographic GWD at 60°S. Each comprises an ensemble of twelve 5-yr transient simulations, yielding a total of 60 yr of simulation for each experiment. Since CMAM has a stratospheric chemistry scheme, it also predicts ozone; we will discuss the impact of the extra GWD on ozone, and its implications, at the end of this section.

The control simulation exhibits the cold-pole bias that is common to nearly all CCMs, namely a too-cold and too-strong winter polar vortex, which breaks down too late in the year (about 2–3 weeks in this case). We demonstrate that these biases are not due to insufficient resolved wave forcing from the troposphere, indicating that they must be due to insufficient (parameterized) GWD in the stratosphere.

In the simulation with extra orographic GWD at 60°S the zonal mean zonal wind and temperature biases in

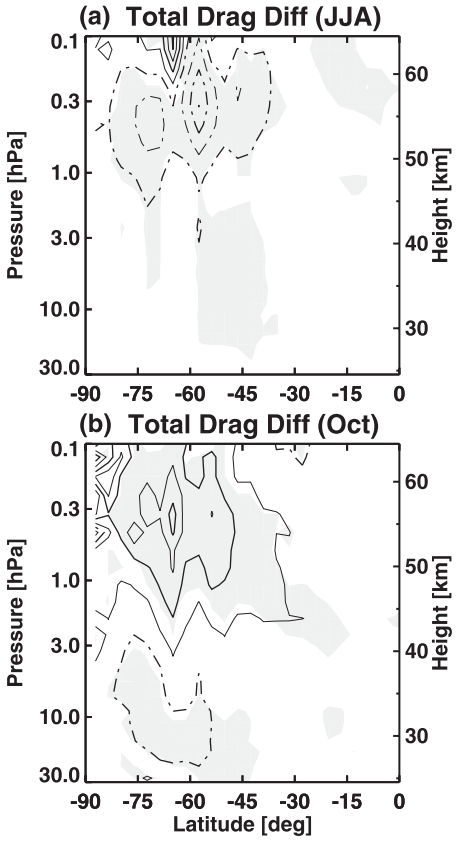

FIG. 15. Difference (GWD60S minus control) in total wave drag for (a) JJA and (b) October. Contour intervals in (a) and (b) are as in Figs. 11f and 12f, respectively. Shading denotes regions where the differences are statistically significant at the 95% level. Total wave drag is the sum of the resolved wave drag and total GWD.

SH winter and early spring are significantly reduced, and the polar vortex breaks down closer to the observed date. The extra GWD also results in changes in resolved wave drag, due to changes in the zonal mean winds. Although we did not perform a zonal wavenumber decomposition, the changes in resolved wave drag at these heights must be due primarily to changes in planetary wave drag. In winter and early spring the extra GWD reduces the zonal mean winds near 60°S, which in turn results in a latitudinal spreading of the resolved wave drag as a result of weaker potential vorticity gradients along the jet axis and a correspondingly less effective waveguide for the vertical propagation of planetary waves. This effect is strongest in winter when the zonal mean winds are strongest. This results in changes in total wave drag (i.e., resolved wave drag plus total GWD) that have the same sign as the total GWD changes at 60°S but are more spread out in latitude. The inclusion of extra GWD also reduces polar-cap temperature biases in late spring and early summer. This is due entirely to the reduction in resolved wave drag as a result of the earlier breakdown of the SH vortex. These findings underscore the importance of considering feedbacks between the extra GWD and the resolved waves when diagnosing the total dynamical response to the extra GWD.

We posed the question whether missing orographic GWD near 60°S is the cause of the stratospheric SH winter and spring zonal wind biases in CCMs. Although we cannot definitively reply in the affirmative, all of our evidence points to missing orographic GWD as the likely cause of these biases. Our primary piece of evidence is the presence of the large negative zonal wind analysis increments in CMAM-DAS near 60°S, which descend in time as the zonal mean westerlies decrease: the fact that the analysis increments follow the descending zonal wind contours and are located at heights below the zero-wind line is highly suggestive of zero-phase-speed gravity waves, such as orographic gravity waves. As to why missing orographic GWD could occur at a latitude band where there is no large-scale topography (only ocean), we suggest two possible reasons. The first is meridional propagation of mountain waves into the jet core, which is situated near 60°S, as discussed by Sato et al. (2009). The second is GWD generated by small isolated islands in the 60°S latitude band, as discussed by Alexander et al. (2009). Neither of these effects is included in current orographic GWD parameterizations. Given our findings, modelers should give serious thought to modifying those parameterizations to account for these effects.

We cannot rule out the possibility that the missing GWD is nonorographic. It is conceivable that the nonorographic GWD schemes in CCMs are all missing an important source of nonorographic gravity waves that would generate drag in the stratosphere near 60°S. However, if that were the case then the CMAM-DAS analysis increments indicate that the missing nonorographic gravity waves would have to be strong enough to occur in the stratosphere near 60°S and not elsewhere, which seems unlikely.

In closing we would like to briefly touch on the impact of the extra GWD on the Antarctic ozone hole. Although impacts on ozone are not the focus of this study, the inclusion of the extra GWD does substantially change the ozone over the SH polar cap in springtime. Figure 17 shows daily polar-cap column ozone for the two experiments. Both have ozone minima in mid-October, but the minimum for the GWD60S experiment is about 40 Dobson units (DU) shallower than for the control experiment. This occurs for two reasons. First, the climatological polar-cap temperatures in the GWD60S experiment are warmer, which results in the

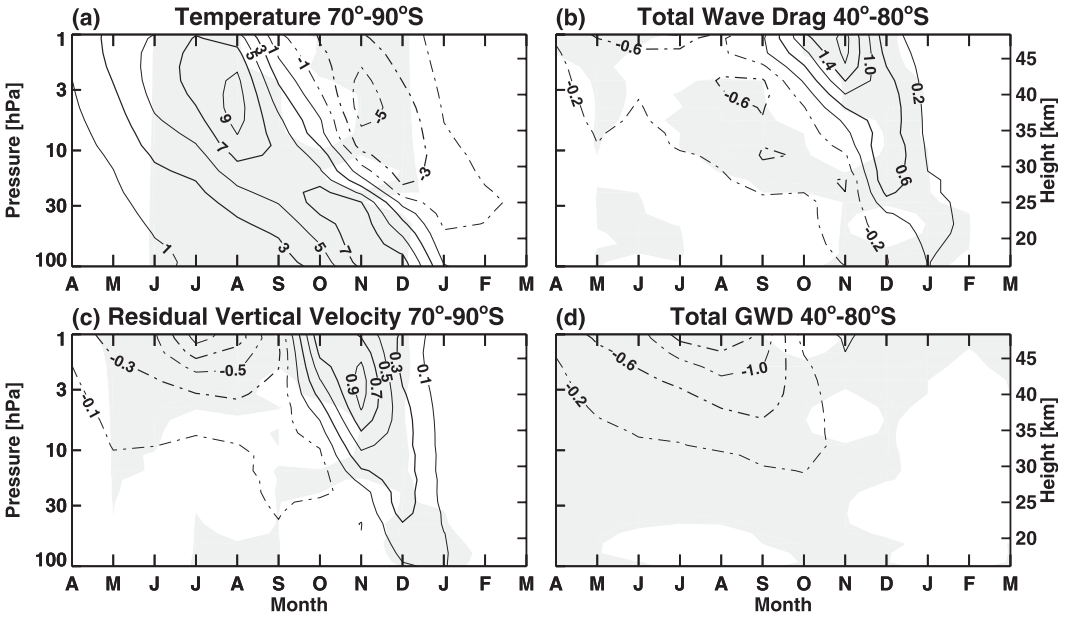

FIG. 16. Differences (GWD60S minus control) of monthly mean (a) temperature, (b) total wave drag, (c) residual vertical velocity, and (d) total GWD. Temperature and vertical velocity are averaged from 70° to 90°S, and wave drag from 40° to 80°S. Contour intervals are 2 K, 0.2 mm s^{-1}, and 0.4 m s^{-1} day^{-1}; negative values are dashed. Shading denotes differences that are statistically significant at the 95% level.

daily temperatures dropping below the threshold for polar stratospheric clouds less often, which in turn results in less heterogeneous chemical ozone loss. Second, a weaker vortex weakens the polar-vortex transport barrier, increasing the flux of ozone from lower latitudes and providing less confinement of ozone-depleted air. Also shown in Fig. 17 are results from the National Institute of Water and Atmosphere (NIWA) combined ozone database (Bodeker et al. 2005), updated version 2.7 "Long-patched" daily data (http://www.bodekerscientific.com/data/total-column-ozone). Since only 5 yr of data are used, there is uncertainty in the mean, although the results do suggest that the ozone hole in the GWD60S experiment is too shallow. This is confirmed by comparing the CMAM results to the three-member ensemble mean monthly mean results from the CMAM CCMVal-1 REF-2 simulations for the same 5-yr period, also shown in Fig. 17. (Since the CCMVal-1 version of CMAM did not have tropospheric chemistry, the global and annual mean column ozone is lower by about 11 DU than for the control experiment. An 11-DU offset has therefore been added to the CCMVal-1 results in Fig. 17.) Not surprisingly, the CCMVal-1 results agree well with the control experiment. The fact that the Antarctic ozone hole in the 20-yr (1980–99) climatology from CMAM CCMVal-1 is somewhat shallower than the corresponding 20-yr climatologies of both the NIWA and satellite-based observations (Fig. 14 of Eyring et al. 2006) confirms that the ozone hole in the GWD60S experiment is indeed too shallow. Thus, reducing the dynamical biases in CMAM by the inclusion of the extra GWD has degraded the simulation of ozone in the Antarctic. This suggests that CCMs that obtain realistic

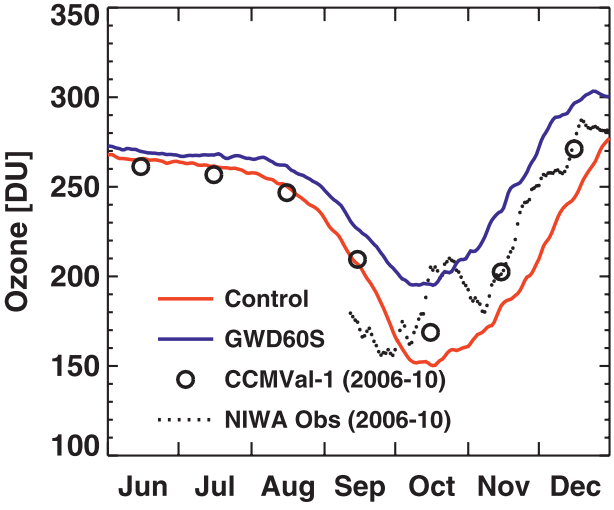

FIG. 17. Daily total column ozone averaged from 70° to 90°S for the control experiment (red), GWD60S experiment (blue), and NIWA "LongPatched" observations for 2006–10 (dotted). Monthly and three-member ensemble mean values from CMAM CCMVal-1 REF-2 simulations from 2006–10 are given by the black circles; an offset of 11 DU has been added to the CCMVal-1 data (see text).

column ozone in the presence of an overly strong and persistent Antarctic polar vortex may be doing so through compensating errors.

Acknowledgments. CM thanks Norm McFarlane, Isla Simpson, Diane Pendlebury, and Andreas Jonsson for helpful discussions, as well as Mike Neish for technical assistance. The authors thank Shuzhan Ren for providing the CMAM-DAS analysis increments and Rolando Garcia and an anonymous reviewer for their constructive reviews. This work was funded by the Canadian Foundation for Climate and Atmospheric Sciences and the Canadian Space Agency.

REFERENCES

Alexander, M. J., S. D. Eckermann, D. Broutman, and J. Ma, 2009: Momentum flux estimates for South Georgia Island mountain waves in the stratosphere observed via satellite. *Geophys. Res. Lett.,* **36,** L12816, doi:10.1029/2009GL038587.

Arblaster, J. M., and G. A. Meehl, 2006: Contributions of external forcings to southern annular mode trends. *J. Climate,* **19,** 2896–2905.

Austin, J., and Coauthors, 2003: Uncertainties and assessments of chemistry-climate models of the stratosphere. *Atmos. Chem. Phys.,* **3,** 1–27.

Black, R. X., and B. A. McDaniel, 2007: Interannual variability in the Southern Hemisphere circulation organized by stratospheric final warming events. *J. Atmos. Sci.,* **64,** 2968–2974.

Bodeker, G. E., H. Shiona, and H. Eskes, 2005: Indicators of Antarctic ozone depletion. *Atmos. Chem. Phys.,* **5,** 2603–2615.

Butchart, N., and Coauthors, 2011: Multimodel climate and variability of the stratosphere. *J. Geophys. Res.,* **116,** D05102, doi:10.1029/2010JD014995.

Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, 457 pp.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.,* **137,** 553–597, doi:10.1002/qj.828.

de Grandpré, J., S. R. Beagley, V. I. Fomichev, E. Griffioen, J. C. McConnell, A. S. Medvedev, and T. G. Shepherd, 2000: Ozone climatology using interactive chemistry: Results from the Canadian Middle Atmosphere Model. *J. Geophys. Res.,* **105,** 26 475–26 491.

Eyring, V., D. E. Kinnison, and T. G. Shepherd, 2005: Overview of planned coupled chemistry–climate simulations to support upcoming ozone and climate assessments. *SPARC Newsletter,* SPARC International Project Office, Toronto, ON, Canada, Vol. 25, 11–17.

——, and Coauthors, 2006: Assessment of temperature, trace species, and ozone in chemistry–climate model simulations of the recent past. *J. Geophys. Res.,* **111,** D22308, doi:10.1029/2006JD007327.

——, T. G. Shepherd, and D. W. Waugh, Eds., 2010: SPARC CCMVal report on the evaluation of chemistry–climate models. SPARC Rep. 5, WCRP-132, WMO/TD-No. 1526, 434 pp. [Available online at http://www.sparc-climate.org/publications/sparc-reports/sparc-report-no5/.]

Garcia, R. R., and B. A. Boville, 1994: "Downward control" of the mean meridional circulation and temperature distribution of the polar winter stratosphere. *J. Atmos. Sci.,* **51,** 2238–2245.

Manzini, E., B. Steil, C. Brühl, M. A. Giorgetta, and K. Krüger, 2003: A new interactive chemistry–climate model: 2. Sensitivity of the middle atmosphere to ozone depletion and increase in greenhouse gases and implications for recent stratospheric cooling. *J. Geophys. Res.,* **108,** 4429, doi:10.1029/2002JD002977.

Matsuno, T., 1970: Vertical propagation of stationary planetary waves in the winter Northern Hemisphere. *J. Atmos. Sci.,* **27,** 871–883.

McFarlane, N. A., 1987: The effect of orographically excited gravity wave drag on the general circulation of the lower stratosphere and troposphere. *J. Atmos. Sci.,* **44,** 1775–1800.

McLandress, C., and N. A. McFarlane, 1993: Interactions between orographic gravity wave drag and forced stationary planetary waves in the winter Northern Hemisphere middle atmosphere. *J. Atmos. Sci.,* **50,** 1966–1990.

——, T. G. Shepherd, J. F. Scinocca, D. A. Plummer, M. Sigmond, A. I. Jonsson, and M. C. Reader, 2011: Separating the dynamical effects of climate change and ozone depletion. Part II: Southern Hemisphere troposphere. *J. Climate,* **24,** 1850–1868.

Morgenstern, O., and Coauthors, 2010: Review of the formulation of present-generation stratospheric chemistry-climate models and associated external forcings. *J. Geophys. Res.,* **115,** D00M02, doi:10.1029/2009JD013728.

Polavarapu, S., S. Ren, A. M. Clayton, D. Sankey, and Y. Rochon, 2004: On the relationship between incremental analysis updating and incremental digital filtering. *Mon. Wea. Rev.,* **132,** 2495–2502.

——, ——, Y. Rochon, D. Sankey, N. Ek, J. Koshyk, and D. Tarasick, 2005a: Data assimilation with the Canadian Middle Atmosphere Model. *Atmos.–Ocean,* **43,** 77–100.

——, T. G. Shepherd, Y. Rochon, and S. Ren, 2005b: Some challenges of middle atmosphere data assimilation. *Quart. J. Roy. Meteor. Soc.,* **131,** 3513–3527.

Ren, S., S. Polavarapu, S. R. Beagley, Y. Nezlin, and Y. J. Rochon, 2011: The impact of gravity wave drag on mesospheric analyses of the 2006 stratospheric major warming. *J. Geophys. Res.,* **116,** D19116, doi:10.1029/2011JD015943.

Sato, K., S. Watanabe, Y. Kawatani, Y. Tomikawa, K. Miyazaki, and M. Takahashi, 2009: On the origins of mesospheric gravity waves. *Geophys. Res. Lett.,* **36,** L19801, doi:10.1029/2009GL039908.

Scinocca, J. F., 2003: An accurate spectral nonorographic gravity wave drag parameterization for general circulation models. *J. Atmos. Sci.,* **60,** 667–682.

——, and N. A. McFarlane, 2000: The parametrization of drag induced by stratified flow over anisotropic orography. *Quart. J. Roy. Meteor. Soc.,* **126,** 2353–2393.

——, ——, M. Lazare, J. Li, and D. Plummer, 2008: The CCCma third generation AGCM and its extension into the middle atmosphere. *Atmos. Chem. Phys.,* **8,** 7055–7074.

Stolarski, R. S., A. R. Douglass, M. Gupta, P. A. Newman, S. Pawson, M. R. Schoeberl, and J. E. Nielsen, 2006: An ozone increase in the Antarctic summer stratosphere: A dynamical response to the ozone hole. *Geophys. Res. Lett.,* **33,** L21805, doi:10.1029/2006GL026820.

Waugh, D. W., W. J. Randel, S. Pawson, P. A. Newman, and E. R. Nash, 1999: Persistence of the lower stratospheric polar vortices. *J. Geophys. Res.,* **104,** 27 191–27 202.